Paleobot.org: establishing open-access online reference collections for archaeobotanical research

Warinner, C ; d'Alpoim Guedes, J ; Goode, D

Abstract: Like other analytic aspects of archaeology, archaeobotany has been growing progressively more quantitative in the past few decades. This may be a sign of the proliferation of increasingly mature and sophisticated methodologies for analyzing botanical data, but associated with the sophistication of quantitative methods is their inherent opacity: the value and applicability of anthropological conclusions drawn from quantitative archaeobotanical data are not only limited by the amount of information that can be extracted from data by sophisticated statistical tools, but also by our ability to draw reasonable anthropological—as opposed to merely statistical—conclusions. Even the words "classification" and "significance" have different meanings in statistics and in anthropology. In this paper, I propose the use of graphical analysis for archaeobotanical data in addition to, or instead of, typical statistical tools like significance tests, variable reduction, and clustering. Applied to data from charred seed assemblages from the ancient Near East, the visual representation of quantitative data has the advantage of handling semi-quantitative data better and being interpretable without reliance on the paradigm of a formal statistical test.

DOI: https://doi.org/10.1007/s00334-011-0282-6

Veget Hist Archaeobot (2011) 20:241–244
DOI 10.1007/s00334-011-0282-6

SHORT COMMUNICATION

Paleobot.org: establishing open-access online reference collections for archaeobotanical research

Christina Warinner · Jade d'Alpoim Guedes · David Goode

Received: 30 September 2010 / Accepted: 22 January 2011 / Published online: 8 February 2011

Abstract Difficulty in accessing high quality reference materials has been a limiting factor in the advancement of archaeobotanical research. However, new developments in online open source content management technology and faster downloading capabilities make high quality and low cost dynamic online curation of archaeobotanical reference images increasingly feasible. We describe the establishment of *Paleobot.org*, an open access online reference collection database for macrobotanical, microbotanical and isotopic data to help standardize and improve the identification of archaeobotanical remains.

Keywords Digital reference collection · Macrobotanical remains · Phytolith · Starch · Pollen · Archaeobotany

Communicated by F. Bittmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00334-011-0282-6) contains supplementary material, which is available to authorized users.

C. Warinner (✉)
Centre for Evolutionary Medicine, Institute of Anatomy, University of Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland
e-mail: christina.warinner@uzh.ch

J. d'Alpoim Guedes
Department of Anthropology, Harvard University, 11 Divinity Ave, Cambridge, MA 02138, USA

D. Goode
Harvard College, Harvard University, University Hall, Cambridge, MA 02138, USA

Introduction

The level of confidence with which one is able to make an accurate taxonomic identification of a given archaeobotanical specimen is closely related to the size and quality of the available reference collection. As a result, access to a suitable reference collection is essential for the positive identification of unknown archaeobotanical remains. As archaeobotanists are now working across the globe, the establishment of large, regionally based reference collections poses increasing challenges. The identification of unknowns using herbaria alone can be costly both in terms of travel and time. In many regions, access to herbaria and identification guides is poor, and where access does exist, many institutions discourage repeated sampling of important type specimens. Online reference collections can greatly lower the cost of establishing a relevant comparative collection, while also reducing over-sampling and redundant destructive analysis of valuable type material.

As an aid to archaeobotanists both in the field and at their home institutions, high-quality online reference collections can facilitate rapid preliminary identification, thereby reducing the number of taxa that will need to be viewed and compared using conventional type collections in the final stages of taxonomic identification. Open-access online reference collections also allow individual researchers to make available images of specimens collected from non-traditional sources, including archaeological sites and modern agricultural fields and markets.

These specimens, which are not typically curated by herbaria, may have particular relevance for archaeobotanists concerned with morphological changes that accompany domestication and regional trait selection.

Finally, the development of new fields in microfossil analysis such as starch grain analysis and phytolith

research, have made new demands on the field. Questions such as variability in starch grain and phytolith morphology are currently under investigation, and access to large online reference collections of modern microfossil images would greatly facilitate the establishment of robust diagnostic criteria for positive taxonomic identification.

Demand for online reference collections and forums

In 2010, Dr. Naomi Miller circulated a questionnaire to archaeobotanists working in the field with the goal of determining the state of current archaeobotanical research and identifying areas in need of improvement (Archaeobotany Questionnaire 2010). The results were presented at the 2010 Society for American Archaeology meetings in a forum entitled, "Quantification and presentation: effective means of presenting plant evidence in Archaeology," organized by Christine Hastorf.

One of the major challenges listed by archaeobotanists in the survey was the lack of identification tools and reference collection images for both specific regions and overall. When asked what would facilitate or enhance their own archaeobotanical research in a practical or intellectual way, respondents listed the development of online identification databases as the single most-desired resource. A strong desire for a website dedicated to the identification of unknown specimens was also expressed. At present, archaeobotanists most frequently solicit assistance with difficult identifications by disseminating images over email list servers, but email is poorly suited for such purposes and is ill-equipped to handle large image files and complicated discussion threads involving multiple individuals. It became clear that a dedicated online source that could host high quality images, as well as curate associated metadata such as field notes, site information, and comments and discussion, would greatly benefit the archaeobotany community.

Current online resources

There are currently a large number of web-based resources available to archaeobotanists specializing in macro and microfossil remains (Table 1 ESM). These resources range from personal research websites of individual analysts to multi-institution searchable photographic collections of particular plant parts or types. While these websites offer many advantages, there are also a number of practical limitations to currently available archaeobotanical online resources. First, with respect to submission, most online resources are closed access. This means that the images available on a particular website are limited to those uploaded by the website's creators.

Aside from *Inside Wood*, none of the current online resources allow other individuals to dynamically upload images from their own reference collections. For individuals interested in making their own reference collection photographs available online, it has been necessary for them to create their own webpage or website, which can be difficult and time-consuming. Another problem is cost. Although most websites are free, some online resources, such as *Discover Seeds* and the *Digital Plant Atlas Project*, require payment for full access to their online collections. For those wishing to create their own website, the costs of programming, site hosting, and server space rental can be prohibitive. An alternative is to use freely available, third party websites, such as Flickr.com, Ning.com, or Wiki Sites, but these free services also greatly limit website flexibility and control.

Herbaria and botanical gardens are increasingly digitizing their collections, but these digitized images are largely limited to scans of botanical sheets. With the exception of the *Seed Herbarium Image Project* (SHIP), a Harvard University initiative that provides online images of seeds housed within the Arnold Arboretum's North American collection, high quality images of seeds or other plant parts which can be useful as identification tools for palaeoethnobotanists are rarely included in digitized herbarium records. While scans of botanical sheets can be helpful for assessing the taxonomic scope of the collection, they are generally not useful for aiding in the identification of macro or microfossils.

In terms of scope, several currently available archaeobotanical websites have a strong regional focus. This poses challenges to researchers engaged in cross-regional research or who work in understudied regions. Additionally, many online collections are biased toward common taxa, and taxonomic coverage is frequently incomplete.

There are also relatively few images of archaeological plants available on current botanical websites. Taphonomic processes such as dessication and carbonization can dramatically alter the appearance of botanical macro and microfossils, and reference collection images of archaeological specimens that have undergone these processes can greatly aid with the identification of unknowns. The Digital Plant Atlas Project published an online *Digital Atlas of Economic Plants* in 2010 (Cappers et al. 2010), and an online *Digital Atlas of Economic Plants in Archaeology* is in preparation, but the images will be limited to archaeobotanical remains excavated from Europe, western Asia and North Africa, by the Deutsches Archäologisches Institut and the Groningen Institute of Archaeology. Geographical coverage of other online images of archaeological plant remains is generally restricted to only a few, mostly Old World, cultural areas.

One major drawback of current resources is that there are very few online forums available for the discussion of

unidentified archaeological botanical remains. Although many researchers currently use listservs to discuss unidentified plant remains, there are no websites that allow individual researchers to upload images of unknowns and solicit assistance with identification.

Finally, current online resources for archaeobotanical research are heavily biased toward macrobotanical remains. There are few online resources available for researchers engaged in the identification of phytoliths and starch microfossils, and there are currently no online databases of plant isotopic data.

Paleobot.org

In order to address the diverse online needs of the archaeobotany community, we have developed *Paleobot.org*, a collaborative, open-access website that allows individual researchers to share reference collection images and data, as well as images of unidentified remains, using a free online platform. The goal of *Paleobot.org* is to bring together a large academic community of archaeobotanists to share data, information, and expertise for the common purpose of improving the identification of archaeobotanical specimens. As with other online resources, the goal of *Paleobot.org* is not to duplicate the function of a herbarium, reference collection specimens, or slides. Rather, the aim of the website is to assist researchers with preliminary identifications, which can then be confirmed with the use of reference collections, literature, or the assistance of a specialist. In other words, *Paleobot.org* is not a replacement for reference collections or plant taxonomists, but rather a way to use them more effectively. *Paleobot.org* was officially launched on July 1, 2010, and the site can be accessed at http://www.paleobot.org. *Paleobot.org* is distinguished by a number of user-friendly features that facilitate scholarly communication and collaborative research, including: open-access viewing and submission, the support of multiple data types, an unidentified forum, assistance with copyright retention, and the ability to create personal research profiles.

Open access: submission and viewing

Open access submission has proven to be a powerful tool for the rapid development of new research and the dissemination of data to both professionals and the public. Examples of successful open access web-databases in the field of biology include *GenBank* (Benson et al. 2008) and Florida State University's *Morphbank* (2010), and the websites *Bonecommons* and *Zooarchaeology.ning.com* offer open-access platforms for individuals engaged in zooarchaeological research (Morris 2009). Current non-open access botanical websites often contain data from either a regional herbarium or an individual's regionally specific collection. Individuals working in areas, for which taxa from two different regions must be consulted, such as the Eurasian steppes, must attempt to access reference collections from widely disparate sources. When such websites contain data from archaeological sites, they are also limited in regional scope and focus. Open access submission is important for allowing data from otherwise underrepresented regions to be accessible to the scholarly community. In addition, open access submission will allow individuals who otherwise would not have the means to develop a website to place their collection online.

Data types

Archaeobotanists work with a wide variety of macro and microbotanical remains and require access to large reference collections to be able to effectively and accurately identify specimens. *Paleobot.org* curates images of the most commonly analyzed types of macro- and microbotanical remains, including macrobotanical remains, pollen, phytoliths and starches, to facilitate this process. In addition to image files, C, N, O, and H isotopic data may also be uploaded. A planned members-only portion of the website will soon allow scholars to share additional documents and information, including identification guides. We anticipate that *Paleobot.org* will play an important role in the development of identification procedures for newer fields such as phytolith research and starch grain analysis. In particular, little is understood of intra species/genus/family variation in starch due to a lack of access to large numbers of reference collection images. We hope that by making large numbers of images available online a better understanding of intra taxa variation will be achieved.

Unidentified forum

All archaeobotanists, at one time or another, encounter a specimen that eludes identification. *Paleobot.org* provides an unidentified forum where researchers can upload images of unidentified specimens and solicit advice from fellow archaeobotanists. Comments and discussion threads are curated with the image for easy reference and may later serve as an archived source of information for similar specimens.

Intellectual property

We acknowledge that reference collections take considerable time and effort to construct. For an open-access online

database to be successful, it is essential that contributors are properly credited for their reference collection images. *Paleobot.org* takes a number of measures to ensure that the authors of these images are given proper citation and acknowledgment for their work. First, all users retain copyright over the personal research images that they upload to the site. This differs from some other sites where the hosting institution, for example a university, controls the copyright of any posted material. Clear copyright and fair use guidelines are posted on the website, and in order to ensure that images placed on *Paleobot.org* cannot be used in presentations or publications without proper image credit, we employ software that automatically embeds copyright information on all images uploaded to the site. Under the terms of fair use, users agree to allow individuals to use content and images from *Paleobot.org* for their own identification purposes and in non-commercial scholarly and educational materials without special permission as long as they include proper photo credit and citation.

Personal research profile

Contributors of images to the website also gain exposure through the creation of a personal research profile on our Contributors page. Although anyone may search the online collections, contributors to the website are required to create a personal research profile before they submit images to the website. Following the creation of a profile, each contributor's photograph, institutional affiliation, contact email address, research website links, and research interests will be featured on the contributors tab of the website. We anticipate that this page will serve as a hub for connecting researchers around the world.

How *Paleobot.org* works

Paleobot.org is deployed using PHP as the backend processing language which creates the pages displayed to the user, and MySQL for the database which provides and stores the updatable content. The site is created using the Drupal framework, an open source software suite that manages most basic content submission and user authentication. Coupled with other open source extensions to this framework, termed modules, many useful behaviours such as location handling, image uploading and content searching can be implemented. Using various extensions, this framework was customized to work as desired for this site. The user-submitted data is stored with meta-data for each specimen as one entity, to which various observations can be attached, such as macrobotanical or phytolith photographs. This allows for varied types of data to be stored within a unified, easily searchable format, and for scientists to subsequently add additional data that is correctly associated with the same specimen.

Conclusions

Paleobot.org is a new, collaborative online resource aimed at facilitating archaeobotanical research. *Paleobot.org* provides free, open-access hosting of reference collection images and data provided by contributing researchers, as well as a dynamic forum for the discussion of unidentified archaeobotanical remains. The goal of *Paleobot.org* is to bring together the archaeobotany community to share data, information, and expertise for the common purpose of improving the identification of archaeobotanical specimens. We invite fellow researchers to submit their reference collections, archaeobotanical and unidentified photographs and data to *Paleobot.org* and to participate in our uniquely collaborative archaeobotanical web-database.

References

Archaeobotany questionnaire (2010) University of Pennsylvania School of Arts and Sciences. http://www.sas.upenn.edu/~nmiller0/AbotQ.pdf. Accessed September 27, 2010

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Res 36:D25–D30

Cappers RTJ, Neef R, Bekker RM (2010) Digital atlas of economic plants, vols 1–3. Barkhuis, Eelde

Morphbank (2010) Florida State University, Department of Scientific Computing, http://www.morphbank.net. Accessed September 25, 2010

Morris J (2009) Zoobook: a zooarchaeology social network. International Council for Archaeozoology Newsletter 10:2